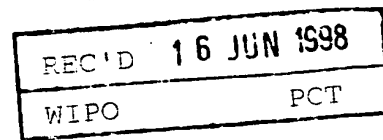




PC/AU98/00380

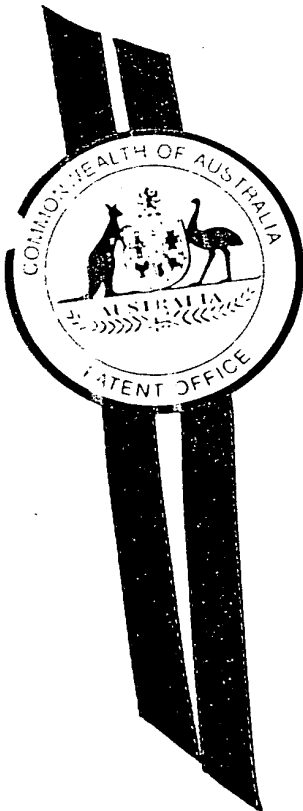


Patent Office
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES,
hereby certify that the annexed is a true copy of the Provisional specification in
connection with Application No. PP 1459 for a patent by THE COUNCIL OF THE
QUEENSLAND INSTITUTE OF MEDICAL RESEARCH filed on 22 January 1998.

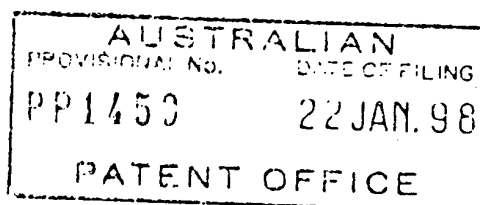
I further certify that the annexed specification is not, as yet, open to public inspection.

PRIORITY DOCUMENT



WITNESS my hand this First
day of June 1998

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND
SALES



The Council of The Queensland Institute of Medical Research

A U S T R A L I A

Patents Act 1990

PROVISIONAL SPECIFICATION

for the invention entitled:

"A novel gene and uses therefore-IIc"

The invention is described in the following statement:

- 1A -

A NOVEL GENE AND USES THEREFOR-IIc**FIELD OF THE INVENTION**

5

The present invention relates generally to a novel human gene and to derivatives and mammalian, animal, insect, nematode, avian and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

10

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description. Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined after the bibliography.

15 BACKGROUND OF THE INVENTION

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis, conventional pharmaceutical
20 preparations as well as gene and protein replacement therapies.

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. On area of particular interest is in the field of heat shock proteins. The *Escherichia coli* heat shock protein
25 DnaJ is the founding member of a family of proteins that are associated with protein folding, protein complex assembly and transit through subcellular components.

Prokaryotic and eukaryotic DnaJ homologues have a modular organisation consisting of a J domain, a glycine-rich spacer, CXXCXGXG repeats and a C-terminal region with no obvious
30 sequence features, as well as additional sequences for protein targeting. The J domain is anticipated to mediate interaction with heat shock 70 proteins (Hsp70) and consists of some 70

amino acids, frequently located at the N-terminus of the protein.

In accordance with the present invention, the inventors have identified a novel gene on chromosome 11q13 which encodes a protein. It is proposed that the protein is a heat shock
5 protein and may have a role in tumour suppression.

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or
10 variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

One aspect of the present invention contemplates an isolated nucleic acid molecule comprising
15 a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a derivative of said heat shock protein or heat shock binding protein.

Another aspect of the present invention is directed to an isolated nucleic acid molecule
20 comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:1;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:2;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence
25 of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

Even yet another aspect of the present invention provides a genetic construct comprising a vector
30 portion and an animal, more particularly a mammalian and even more particularly a human *mcl8* gene portion, which *mcl8* gene portion is capable of encoding an MCG18 polypeptide

- 3 -

or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence
5 of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

10 Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

15 Another aspect of the present invention contemplates a method for detecting MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

20

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the nucleotide sequence and corresponding amino acid sequence of *mcg18*.

25

Figure 2 is a representation showing that MCG18 has partial homology to *E. coli* DnaJ.

Figure 3 is a representation showing that MCG18 has homology to two *Caenorhabditis elegans* proteins.

30

Figure 4 is a representation showing that MCG18 has homology to a *Saccharomyces pombe*

protein.

Figure 5 is a representation showing homology of MCG18 to a *Drosophila virilis* protein.

5 **Figure 6** is a representation showing homology of MCG18 to human DnaJ proteins HDJ-2/HSDJ, HDJ-1/HSP40 and HSJ1.

Figure 7 is a representation of the nucleotide and corresponding amino acid sequence of murine *mcg18*.

10

Figure 8 is a representation of homology between human and murine MCG18.

Figure 9 depicts nucleotide sequences corresponding to the 5' untranslated region of human *mcg18*.

15

Figure 10 depicts a Northern blot showing expression of *mcg18* transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain 15 μ g RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

20

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Accordingly, one aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an
25 amino acid sequence having homology to a heat shock protein or a derivative of said heat shock protein or a heat shock-binding protein.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

30

- (i) a nucleotide sequence set forth in SEQ ID NO:1;

- 5 -

- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:2;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the
- 5 nucleotide sequence set forth in (i), (ii) or (iii).

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

- 10 Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and
- 15 from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

20

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

- 25 Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

The present invention extends to nucleic acid molecules with percentage similarities of

30 approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

The nucleic acid molecule of the present invention is hereinafter referred to as constituting the "*mcg18*" gene. The protein encoded by *mcg18* is referred to herein as "MCG18".

The present invention extends to the naturally occurring genomic *mcg18* nucleotide sequence
5 or corresponding cDNA sequence or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG18 or the corresponding genetic sequence. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to *mcg18*. "Additions" to the amino acid or nucleotide
10 sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG18" or "*mcg18*" includes references to all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG18.

The *mcg18* of the present invention is particularly exemplified herein from humans and in
15 particular from human chromosome 11q13.

The present invention extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), birds (eg. chickens, ducks, geese,
20 parrot), insects, nematodes, eukaryotic microorganism and captive wild animals (eg. deer, foxes, kangaroos). Reference herein to *mcg18* or MCG18 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic
25 acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic
30 molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or

- 7 -

both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

- 5 Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human *mcg18* gene portion, which *mcg18* gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.
- 10 Preferably, the *mcg18* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg18* gene portion in an appropriate cell.

In addition, the *mcg18* gene portion of the genetic construct may comprise all or part of the gene
15 fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

20

It is proposed in accordance with the present invention that MCG18 is a transcription factor involved in protein folding, protein complex assembly and transit through subcellular compartments. MCG18 may also have a role in tumour suppression. Thus mutations in *mcg18* may result in the development of or a propensity to develop various types of cancer.

25

A deletion or aberration in the *mcg18* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may
30 be determined by assaying for aberrations in the parents and/or proband of the subject under investigation.

- 8 -

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said *mcg18* wherein the presence of such a nucleotide
5 substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

The nucleotide substitutions, additions or deletions may be detected by any convenient means including nucleotide sequencing, restriction fragment length polymorphism (RFLP), polymerase
10 chain reaction (PCR), oligonucleotide hybridization and single stranded conformation polymorphism analysis (SSCP) amongst many others. An aberration includes modifications to existing nucleotides such as to modify glycosylation signals amongst other effects.

In an alternative method, aberrations in the *mcg18* gene are detected by screening for mutations
15 in MCG18.

A mutation in MCG18 may be a single or multiple amino acid substitution, addition and/or deletion. The mutation in *mcg18* may also result in either no translation product being produced or a product in truncated form. A mutation may also be an altered glycosylation pattern or the
20 introduction of side chain modifications to amino acid residues.

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the
25 presence of such a mutation is indicative of or a propensity to develop said condition.

A particularly convenient means of detecting a mutation in MCG18 is by use of antibodies.

Accordingly another aspect of the present invention is directed to antibodies to MCG18 and its
30 derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG18 or may be specifically raised to MCG18 or derivatives

- 9 -

thereof. In the case of the latter, MCG18 or its derivatives may first need to be associated with a carrier molecule. The antibodies to MCG18 of the present invention are particularly useful as diagnostic agents.

5 For example, antibodies to MCG18 and its derivatives can be used to screen for wild-type MCG18 or for mutated MCG18 molecules. The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG18 levels or the presence of wild-type MCG18 may be
10 important for diagnosis of certain cancers or a predisposition for development of cancers or for monitoring certain therapeutic protocols.

As stated above antibodies to MCG18 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention
15 extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

For example, specific antibodies can be used to screen for wild-type MCG18 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for
20 example, as a means for screening for levels of MCG18 in a cell extract or other biological fluid or purifying MCG18 made by recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

25 It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG4 or to a
30 specific mutant phenotype or to a deleted or otherwise altered region.

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG18 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal or bird
5 with an effective amount of MCG18 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

10

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques
15 which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and
20 under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

25 The presence of MCG18 may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4, 424,279 and 4,018,653. These includes both single-site and two-site or "sandwich" assays of the non-competitive types, as well as in traditional competitive binding assays. These assays also include direct binding of a labelled
30 antibody to a target.

- 11 -

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into
5 contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is
10 determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the
15 art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG18 including cell extract or, tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

20 In the typical forward sandwich assay, a first antibody having specificity for the MCG18 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates,
25 or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes) and under suitable conditions (e.g. 25°C) to allow binding of any
30 subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the

- 12 -

happen. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of

- 13 -

agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination
5 with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate
10 wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

15 As stated above, MCG18 is proposed to have a role in tumour suppression. Accordingly, it is further proposed in accordance with the present invention to use recombinant MCG18 in pharmaceutical preparations for treating arresting or otherwise ameliorating the effects of certain cancers.

20 Accordingly, another aspect of the present invention contemplates a method for treating, arresting or otherwise ameliorating the effects of a cancer in an animal or bird, said method comprising administering to said animal or bird an effective amount of MCG18 or a functional derivative thereof for a time and under conditions sufficient to treat, arrest or otherwise ameliorate the effects of said cancer.

25

The present invention, therefore, contemplates a pharmaceutical composition comprising MCG18 or a derivative thereof or a modulator of *mcg18* expression or MCG18 activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to hereinafter as the "active ingredients". The active ingredients may also include anti-cancer
30 agents or agents which facilitate actions of MCG18.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion or may be in the form of a cream or other form suitable for topical application. It must be stable under the conditions of manufacture and storage and must
5 be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of
10 dispersion and by the use of surfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying
15 absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating
20 the various sterilized active ingredient into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

25

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be
30 incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations

should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions is such that a suitable dosage will be obtained. Preferred compositions or
5 preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 μ g and 2000 mg of active compound.

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter: A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium
10 phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to
15 otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In
20 addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

The present invention also extends to forms suitable for topical application such as creams, lotions and gels.

25

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active
30 ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

- 16 -

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 μ g to about 2000 mg. Expressed in proportions, the active compound is generally present in from about 0.5 μ g to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating *mcl8* expression or MCG18 activity. The vector may, for example, be a viral vector.

As stated above, the present invention further contemplates a range of derivatives of MCG18. Derivatives include fragments, parts, portions, mutants, homologues and analogues of the MCG18 polypeptide and corresponding genetic sequence. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to the genetic sequence encoding MCG18. "Additions" to amino acid sequences or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG18" includes reference to all derivatives thereof including functional derivatives or MCG18

immunologically interactive derivatives.

Analogues of MCG18 contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, 5 polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde 10 followed by reduction with NaBH_4 ; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with NaBH_4 .

15

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation 20 followed by subsequent derivitisation, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted 25 maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or 30 alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form

- 18 -

a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

5

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of
10 amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 1.

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with $n=1$ to $n=6$, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually
15 contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_ϵ -methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an
20 amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

The present invention further contemplates chemical analogues of MCG18 capable of acting as antagonists or agonists of MCG18 or which can act as functional analogues of MCG18.
25 Chemical analogues may not necessarily be derived from MCG18 but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of MCG18. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

30 The identification of MCG18 permits the generation of a range of therapeutic molecules capable of modulating expression of MCG18 or modulating the activity of MCG18. Modulators

- 19 -

contemplated by the present invention includes agonists and antagonists of MCG18 expression. Antagonists of MCG18 expression include antisense molecules, ribozymes and co-suppression molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of MCG18 include molecules which overcome any negative
5 regulatory mechanism. Antagonists of MCG18 include antibodies and inhibitor peptide fragments.

TABLE 1

	Non-conventional amino acid	Code	Non-conventional amino acid	Code
5				
	α -aminobutyric acid	Abu	L-N-methylalanine	Nmala
	α -amino- α -methylbutyrate	Mgab	L-N-methylarginine	Nmarg
	aminocyclopropane- carboxylate	Cpro	L-N-methylasparagine	Nmasn
			L-N-methylaspartic acid	Nmasp
10	aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
	aminonorbornyl- carboxylate	Norb	L-N-methylglutamine	Nmgln
			L-N-methylglutamic acid	Nmglu
	cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
	cyclopentylalanine	Cpen	L-N-methylisoleucine	Nmile
15	D-alanine	Dal	L-N-methylleucine	Nmleu
	D-arginine	Darg	L-N-methyllysine	Nmlys
	D-aspartic acid	Das	L-N-methylmethionine	Nmmet
	D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
	D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
20	D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
	D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
	D-isoleucine	Dile	L-N-methylproline	Nmpro
	D-leucine	Dleu	L-N-methylserine	Nmser
	D-lysine	Dlys	L-N-methylthreonine	Nmthr
25	D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
	D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
	D-phenylalanine	Dphe	L-N-methylvaline	Nmval
	D-proline	Dpro	L-N-methylethylglycine	Nmetg
	D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
30	D-threonine	Dthr	L-norleucine	Nle
	D-tryptophan	Dtrp	L-norvaline	Nva

- 21 -

	D-tyrosine	Dtyr	α -methyl-aminoisobutyrate	Maib
	D-valine	Dval	α -methyl- γ -aminobutyrate	Mgab
	D- α -methylalanine	Dmala	α -methylcyclohexylalanine	Mchexa
	D- α -methylarginine	Dmarg	α -methylcyclopentylalanine	Mcpen
5	D- α -methylasparagine	Dmasn	α -methyl- α -naphthylalanine	Manap
	D- α -methylaspartate	Dmasp	α -methylpenicillamine	Mpen
	D- α -methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D- α -methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D- α -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D- α -methylisoleucine	Dmile	N-amino- α -methylbutyrate	Nmaabu
	D- α -methylleucine	Dmleu	α -naphthylalanine	Anap
	D- α -methyllysine	Dmlys	N-benzylglycine	Nphe
	D- α -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D- α -methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D- α -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D- α -methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D- α -methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D- α -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D- α -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D- α -methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D- α -methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl)glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

D-N-methyllysine	Dnmlys	N-methyl- γ -aminobutyrate	Nmgabu
N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
5 N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
D-N-methyltyrosine	Dnmtyr	N-methyla-naphthylalanine	Nmanap
10 D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
γ -aminobutyric acid	Gabu	N-(<i>p</i> -hydroxyphenyl)glycine	Nhtyr
L- <i>t</i> -butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
L-ethylglycine	Etg	penicillamine	Pen
L-homophenylalanine	Hphe	L- α -methylalanine	Mala
15 L- α -methylarginine	Marg	L- α -methylasparagine	Masn
L- α -methylaspartate	Masp	L- α -methyl- <i>t</i> -butylglycine	Mtbug
L- α -methylcysteine	Mcys	L-methylethylglycine	Metg
L- α -methylglutamine	Mgln	L- α -methylglutamate	Mglu
L- α -methylhistidine	Mhis	L- α -methylhomophenylalanine	Mhphe
20 L- α -methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
L- α -methylleucine	Mleu	L- α -methyllysine	Mlys
L- α -methylmethionine	Mmet	L- α -methylnorleucine	Mnle
L- α -methylnorvaline	Mnva	L- α -methylornithine	Morn
L- α -methylphenylalanine	Mphe	L- α -methylproline	Mpro
25 L- α -methylserine	Mser	L- α -methylthreonine	Mthr
L- α -methyltryptophan	Mtrp	L- α -methyltyrosine	Mtyr

L- α -methylvaline	Mval	L-N-methylhomophenylalanine	Nmhphe
N-(N-(2,2-diphenylethyl)	Nnbhm	N-(N-(3,3-diphenylpropyl)	Nnbhe
carbamylmethyl)glycine		carbamylmethyl)glycine	
1-carboxy-1-(2,2-diphenyl-	Nmbc		
5 ethylamino)cyclopropane			

These types of modifications may be important to stabilise MCG18 if administered to an individual or for use as a diagnostic reagent.

10

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

15

Another embodiment of the present invention contemplates a method for modulating expression of MCG18 in a human, said method comprising contacting the *mcg18* gene encoding MCG18 with an effective amount of a modulator of *mcg18* expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression
20 of *mcg18*. For example, a nucleic acid molecule encoding MCG18 or a derivative thereof may be introduced into a cell to facilitate protection of that cell from becoming cancerous.

Another aspect of the present invention contemplates a method of modulating activity of MCG18 in a human, said method comprising administering to said mammal a modulating
25 effective amount of a molecule for a time and under conditions sufficient to increase or decrease MCG18 activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative of MCG18 or a chemical analogue or truncation mutant of MCG18.

The present invention is further described with reference to the following non-limiting
30 Examples.

- 24 -

EXAMPLE 1

A human gene was identified from chromosome 11q13 that encodes a new member of the DnaJ family of proteins (designated MCG18). This gene (*mcg18*) is expressed as an ~1.4kb mRNA (Fig. 10) and is predicted to encode a 241 amino acid product (Fig. 1).

EXAMPLE 2

MCG18 has partial homology to *E. coli* dnaJ and other human DnaJ family members in that it contains the J domain (Fig. 2).

EXAMPLE 3

MCG18 has greatest homology to functionally undefined proteins from *C. elegans* (Fig. 3) and *S. pombe* (Fig. 4) that also feature the J domain but maintain sequence similarity through the central and C-terminal regions of the proteins.

EXAMPLE 4

The J domain is proposed to mediate interaction with heat shock protein (Hsp70) and consist of some 70 amino acids, frequently located at the N-terminus of the protein. One of these proteins, tumorous imaginal discs (Tid58) from *Drosophila virilis* (Fig. 5) functions as a tumour suppressor.

EXAMPLE 5

A comparison of homology between MCG18 and human DnaJ proteins HDJ-2/H5DJ, HDJ-1/HSP40 and HSJ1 is shown in Fig. 6.

EXAMPLE 6

During the sequence characterisation of the *VRF/VEGFB* promoter region on cosmid CLGW4 [Grimmond *et al*, 1996], which maps to chromosome 11q13 the inventors identified a sequence
5 that exactly matched numerous human and mouse expressed sequence tags (ESTs) in the EST database from a gene which we designated *mcg18*. EST clones for human (GenBank accession number T69741, clone 108172; accession number H40901, clone 177008) and mouse *mcg18* (accession number W34884, clone 350966; accession number W64183, clone 385535) were obtained from Genome Systems Inc. and sequenced with the gene-specific primers shown in
10 Table 2. The EST clones listed in Table 3 were also utilised in generating the full-length coding sequence for human (Figure 1) and mouse (Figure 7) *mcg18*. The EST database also contained *mcg18* cDNA entries that were alternately (or partially) spliced, and in order to understand their ability to encode new polypeptides, the gene structure of *mcg18* was determined by sequencing human and mouse genomic templates with gene-specific primers.

15

Genomic fragments containing the human [Grimmond *et al*, 1996] and murine genes [Townson *et al*, 1996] have been reported previously. Cosmid CLGW4 contains the entire human gene and λ 121 contains the entire mouse gene, as determined by direct sequencing of the templates with the oligonucleotides listed in Table 2. Plasmids containing sub-fragments of λ 121 and
20 cosmid CLGW4 were prepared using plasmid purification kits (Qiagen) and sequenced as described previously [Grimmond *et al*, 1996; Townson *et al*, 1996] using primers designed against cDNA and genomic sequences. The BLAST suite of programs [Altschul *et al*, 1990] was used to compare the sequence data against the nucleotide and protein databases at the National Center for Biotechnology Information (<http://www.ncbi.nih.gov.nlm>). The sequence
25 data were compiled using MacVector 4.2.1 software (IBI-Kodak). ClustalW sequence alignments [Thompson *et al*, 1994] were conducted using the Australian National Genome Information Service computer faculty at the University of Sydney, Australia.

The cDNA sequence of human *mcg18* (Figure 1) was translated in all possible reading frames
30 and compared to the GenBank non-redundant protein database using the program BLASTX [Altschul *et al*, 1990] and the coding region was identified on the basis of showing homology

to the DnaJ family of proteins (Figure 2). The DnaJ domain is encoded within the longest open reading frame and the assigned initiation codon is preceded by an in-frame stop codon (Figure 9). Similar database search results were obtained for the mouse *mcg18* cDNA, and the alignment of human and mouse protein sequences is shown in Figure 8. MCG18 has greatest
5 homology to gene products from *C. elegans* (Figure 3) and *S. pombe* (Figure 4). Although it shares a similar J-domain, MCG18 does not contain other domains described for the tumour suppressor gene from *D. virilis* (Figure 5), nor is it a homologue of other reported human J-domain-containing proteins (Figure 6).

10 To determine the expression pattern of *mcg18*, 15µg of total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer using 20 x SSC (Sambrook *et al.*, 1986). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (³²P-dCTP) cDNA probe (Church and Gilbert, 1984) for
15 *mcg18*. After washes in 0.1 x SSC/0.1% SDS for 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that *mcg18* is expressed as a 1.4kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 10).

20 Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or
25 features.

- 27 -

TABLE 2
***mcg18*-SPECIFIC OLIGONUCLEOTIDES**

	<u>name</u>	<u>sequence 5' to 3'</u>
5	HVESTF	AGC GGG CCA GGC CCC TTC
	HV195F	CAT CCT GGT CCA ATG CGC TC
	HV387F2	GCA CTG AGG AAG TTA AAC GAG C
	HV408R	GCT CGT TTA ACT TCC TCA GTG C
	EXON1REV	GCT CAG CTC CAC AAA GCG GCT
	HVEST426F	ACC AGC TCC GCT CAG GTA G
10	HVEST623R	TCC AGG AGC TGT GTG TTT GG
	SGVESTF3	CCA GTT TCA CAG CGT GAG G
	HVEST631R	CAG CAT GAG GAG GAG GCA G

15

TABLE 3
EST CLONE SEQUENCES USED TO GENERATED HUMAN AND MOUSE
***mcg18* cDNA SEQUENCE COMPOSITES**

<u>EST clone number</u>	<u>organism</u>	<u>GenBank accession number</u>
1g2815	human	D45683
001-T2-18	human	F17225
273748	human	N37043
177008	human	H40901 and H40939
258011	human	N30776
276887	human	N44004
108172	human	T69741
307529	human	W21083 and W32579
342027	human	W60283
354288	mouse	W44038
350966	mouse	W348844
426261	mouse	AA002868
368185	mouse	W53911
385535	mouse	W64183
404472	mouse	W82959
406437	mouse	W83482

BIBLIOGRAPHY

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *J. Mol. Biol.* 215: 403-410.
2. Church, G., and Gilber, W. (1984) *Proc. Natl. Acad. Sci. USA* 18: 1991-1995.
3. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G (1996) *Genome Res.* 6: 124-131.
4. Thompson, J. D., Higgins, D.G., and Gibson, T.J. (1994) *Nucleic Acids Res.* 22: 4673-4680.
5. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun.* 220: 922-928.

- 29 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: The Council of The Queensland Institute of Medical Research
- (ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
- (iii) NUMBER OF SEQUENCES: 2
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: DAVIES COLLISON CAVE
 - (B) STREET: 1 LITTLE COLLINS STREET
 - (C) CITY: MELBOURNE
 - (D) STATE: VICTORIA
 - (E) COUNTRY: AUSTRALIA
 - (F) ZIP: 3000
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: AUSTRALIAN PROVISIONAL
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: HUGHES, DR E JOHN L
 - (C) REFERENCE/DOCKET NUMBER: EJH/AF
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: +61 3 9254 2777
 - (B) TELEFAX: +61 3 9254 2770
 - (C) TELEX: AA 31787

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 832 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA

- 30 -

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 11..733

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCCCGCCGCC	ATG	CCG	CCC	TTA	CTG	CCC	CTG	CGC	CTG	TGC	CGG	CTG	TGG			49
	Met	Pro	Pro	Leu	Leu	Pro	Leu	Arg	Leu	Cys	Arg	Leu	Trp			
	1				5					10						
CCC	CGC	AAC	CCT	CCC	TCC	CGG	CTC	CTC	GGA	GCG	GCC	GCC	GGG	CAG	CGG	97
Pro	Arg	Asn	Pro	Pro	Ser	Arg	Leu	Leu	Gly	Ala	Ala	Ala	Gly	Gln	Arg	
	15					20					25					
TCC	AGA	CCC	AGT	ACT	TAT	TAT	GAA	CTG	TTG	GGG	GTG	CAT	CCT	GGT	GCC	145
Ser	Arg	Pro	Ser	Thr	Tyr	Tyr	Glu	Leu	Leu	Gly	Val	His	Pro	Gly	Ala	
	30				35					40					45	
AGC	ACT	GAG	GAA	GTT	AAA	CGA	GCT	TTC	TTC	TCC	AAG	TCC	AAA	GAG	CTG	193
Ser	Thr	Glu	Glu	Val	Lys	Arg	Ala	Phe	Phe	Ser	Lys	Ser	Lys	Glu	Leu	
				50					55					60		
CAC	CCA	GAC	CGG	GAC	CCT	GGG	AAC	CCA	AGC	CTG	CAC	AGC	CGC	TTT	GTG	241
His	Pro	Asp	Arg	Asp	Pro	Gly	Asn	Pro	Ser	Leu	His	Ser	Arg	Phe	Val	
			65					70					75			
GAG	CTG	AGC	GAG	GCA	TAC	CGT	GTG	CTC	AGC	CGT	GAG	CAG	AGC	CGC	CGC	289
Glu	Leu	Ser	Glu	Ala	Tyr	Arg	Val	Leu	Ser	Arg	Glu	Gln	Ser	Arg	Arg	
		80					85					90				
AGC	TAT	GAT	GAC	CAG	CTC	CGC	TCA	GGT	AGT	CCC	CCA	AAG	TCT	CCA	CGA	337
Ser	Tyr	Asp	Asp	Gln	Leu	Arg	Ser	Gly	Ser	Pro	Pro	Lys	Ser	Pro	Arg	
	95					100					105					
ACC	ACA	GTC	CAT	GAC	AAG	TCT	GCC	CAC	CAA	ACA	CAC	AGC	TCC	TGG	ACA	385
Thr	Thr	Val	His	Asp	Lys	Ser	Ala	His	Gln	Thr	His	Ser	Ser	Trp	Thr	
	110				115					120					125	
CCC	CCC	AAC	GCA	CAG	TAC	TGG	TCC	CAG	TTT	CAC	AGC	GTG	AGG	CCA	CAG	433
Pro	Pro	Asn	Ala	Gln	Tyr	Trp	Ser	Gln	Phe	His	Ser	Val	Arg	Pro	Gln	
				130					135					140		
GGG	CCC	CAG	TTG	AGG	CAG	CAG	CAA	CAC	AAA	CAA	AAC	AAA	CAA	GTG	CTG	481
Gly	Pro	Gln	Leu	Arg	Gln	Gln	Gln	His	Lys	Gln	Asn	Lys	Gln	Val	Leu	
			145					150				155				
GGG	TAC	TGC	CTC	CTC	CTC	ATG	CTG	GCG	GGC	ATG	GGC	CTG	CAC	TAC	ATT	529
Gly	Tyr	Cys	Leu	Leu	Leu	Met	Leu	Ala	Gly	Met	Gly	Leu	His	Tyr	Ile	
		160					165					170				
GCC	TTC	AGG	AAG	GTG	AAG	CAG	ATG	CAC	CTT	AAC	TTC	ATG	GAT	GAA	AAG	577
Ala	Phe	Arg	Lys	Val	Lys	Gln	Met	His	Leu	Asn	Phe	Met	Asp	Glu	Lys	
	175					180					185					
GAT	CGG	ATC	ATC	ACA	GCC	TTC	TAC	AAC	GAA	GCC	CGG	GCA	CGG	GCC	AGG	625
Asp	Arg	Ile	Ile	Thr	Ala	Phe	Tyr	Asn	Glu	Ala	Arg	Ala	Arg	Ala	Arg	
					195				200						205	
GCC	AAC	AGA	GGC	ATC	CTT	CAG	CAG	GAG	CGA	CAA	CGG	CTA	GGG	CAG	CGG	673
Ala	Asn	Arg														

- 31 -

GGC GCC GGC CCC TGA GGGGCTC ACCTGGATGG GGCCTGCAGT GCGTTCCCGC
 Gly Ala Gly Pro *
 240

773

TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTGCGAG

832

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 241 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Pro	Leu	Leu	Pro	Leu	Arg	Leu	Cys	Arg	Leu	Trp	Pro	Arg	Asn
1				5					10					15	
Pro	Pro	Ser	Arg	Leu	Leu	Gly	Ala	Ala	Ala	Gly	Gln	Arg	Ser	Arg	Pro
			20					25					30		
Ser	Thr	Tyr	Tyr	Glu	Leu	Leu	Gly	Val	His	Pro	Gly	Ala	Ser	Thr	Glu
		35					40					45			
Glu	Val	Lys	Arg	Ala	Phe	Phe	Ser	Lys	Ser	Lys	Glu	Leu	His	Pro	Asp
	50					55					60				
Arg	Asp	Pro	Gly	Asn	Pro	Ser	Leu	His	Ser	Arg	Phe	Val	Glu	Leu	Ser
65					70					75					80
Glu	Ala	Tyr	Arg	Val	Leu	Ser	Arg	Glu	Gln	Ser	Arg	Arg	Ser	Tyr	Asp
				85					90					95	
Asp	Gln	Leu	Arg	Ser	Gly	Ser	Pro	Pro	Lys	Ser	Pro	Arg	Thr	Thr	Val
			100					105					110		
His	Asp	Lys	Ser	Ala	His	Gln	Thr	His	Ser	Ser	Trp	Thr	Pro	Pro	Asn
		115					120					125			
Ala	Gln	Tyr	Trp	Ser	Gln	Phe	His	Ser	Val	Arg	Pro	Gln	Gly	Pro	Gln
	130					135					140				
Leu	Arg	Gln	Gln	Gln	His	Lys	Gln	Asn	Lys	Gln	Val	Leu	Gly	Tyr	Cys
145					150					155					160
Leu	Leu	Leu	Met	Leu	Ala	Gly	Met	Gly	Leu	His	Tyr	Ile	Ala	Phe	Arg
				165					170					175	
Lys	Val	Lys	Gln	Met	His	Leu	Asn	Phe	Met	Asp	Glu	Lys	Asp	Arg	Ile
			180					185					190		
Ile	Thr	Ala	Phe	Tyr	Asn	Glu	Ala	Arg	Ala	Arg	Ala	Arg	Ala	Asn	Arg
		195					200					205			

- 32 -

Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg Gln Pro Pro
210 215 220
Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg Gly Ala Gly
225 230 235 240
Pro

DATED this 22nd day of January, 1998

The Council of The Queensland Institute of Medical Research

By DAVIES COLLISON CAVE

Patent Attorneys for the Applicants

FIGURE 1

GCCCCGCCGCC	ATG	CCG	CCC	TTA	CTG	CCC	CTG	CGC	CTG	TGC	CGG	CTG	TGG	49		
	Met	Pro	Pro	Leu	Leu	Pro	Leu	Arg	Leu	Cys	Arg	Leu	Trp			
	1				5					10						
CCC	CGC	AAC	CCT	CCC	TCC	CGG	CTC	CTC	GGA	GCG	GCC	GCC	GGG	CAG	CGG	97
Pro	Arg	Asn	Pro	Pro	Ser	Arg	Leu	Leu	Gly	Ala	Ala	Ala	Gly	Gln	Arg	
	15					20					25					
TCC	AGA	CCC	AGT	ACT	TAT	TAT	GAA	CTG	TTG	GGG	GTG	CAT	CCT	GGT	GCC	145
Ser	Arg	Pro	Ser	Thr	Tyr	Tyr	Glu	Leu	Leu	Gly	Val	His	Pro	Gly	Ala	
	30				35					40					45	
AGC	ACT	GAG	GAA	GTT	AAA	CGA	GCT	TTC	TTC	TCC	AAG	TCC	AAA	GAG	CTG	193
Ser	Thr	Glu	Glu	Val	Lys	Arg	Ala	Phe	Phe	Ser	Lys	Ser	Lys	Glu	Leu	
				50					55					60		
CAC	CCA	GAC	CGG	GAC	CCT	GGG	AAC	CCA	AGC	CTG	CAC	AGC	CGC	TTT	GTG	241
His	Pro	Asp	Arg	Asp	Pro	Gly	Asn	Pro	Ser	Leu	His	Ser	Arg	Phe	Val	
			65					70					75			
GAG	CTG	AGC	GAG	GCA	TAC	CGT	GTG	CTC	AGC	CGT	GAG	CAG	AGC	CGC	CGC	289
Glu	Leu	Ser	Glu	Ala	Tyr	Arg	Val	Leu	Ser	Arg	Glu	Gln	Ser	Arg	Arg	
		80					85					90				
AGC	TAT	GAT	GAC	CAG	CTC	CGC	TCA	GGT	AGT	CCC	CCA	AAG	TCT	CCA	CGA	337
Ser	Tyr	Asp	Asp	Gln	Leu	Arg	Ser	Gly	Ser	Pro	Pro	Lys	Ser	Pro	Arg	
	95					100					105					
ACC	ACA	GTC	CAT	GAC	AAG	TCT	GCC	CAC	CAA	ACA	CAC	AGC	TCC	TGG	ACA	385
Thr	Thr	Val	His	Asp	Lys	Ser	Ala	His	Gln	Thr	His	Ser	Ser	Trp	Thr	
	110				115					120					125	
CCC	CCC	AAC	GCA	CAG	TAC	TGG	TCC	CAG	TTT	CAC	AGC	GTG	AGG	CCA	CAG	433
Pro	Pro	Asn	Ala	Gln	Tyr	Trp	Ser	Gln	Phe	His	Ser	Val	Arg	Pro	Gln	
				130					135					140		
GGG	CCC	CAG	TTG	AGG	CAG	CAG	CAA	CAC	AAA	CAA	AAC	AAA	CAA	GTG	CTG	481
Gly	Pro	Gln	Leu	Arg	Gln	Gln	Gln	His	Lys	Gln	Asn	Lys	Gln	Val	Leu	
			145					150					155			
GGG	TAC	TGC	CTC	CTC	CTC	ATG	CTG	GCG	GGC	ATG	GGC	CTG	CAC	TAC	ATT	529
Gly	Tyr	Cys	Leu	Leu	Leu	Met	Leu	Ala	Gly	Met	Gly	Leu	His	Tyr	Ile	
		160					165					170				
GCC	TTC	AGG	AAG	GTG	AAG	CAG	ATG	CAC	CTT	AAC	TTC	ATG	GAT	GAA	AAG	577
Ala	Phe	Arg	Lys	Val	Lys	Gln	Met	His	Leu	Asn	Phe	Met	Asp	Glu	Lys	
	175					180					185					
GAT	CGG	ATC	ATC	ACA	GCC	TTC	TAC	AAC	GAA	GCC	CGG	GCA	CGG	GCC	AGG	625

Figure 1 (continued)

Asp Arg Ile Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg	
190 195 200 205	
GCC AAC AGA GGC ATC CTT CAG CAG GAG CGA CAA CGG CTA GGG CAG CGG	673
Ala Asn Arg Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg	
210 215 220	
CAG CCG CCA CCA TCC GAG CCA ACC CAA GGC CCC GAG ATC GTG CCC CGG	721
Gln Pro Pro Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg	
225 230 235	
GGC GCC GGC CCC TGA GGGGCTC ACCTGGATGG GGCCTGCAGT GCGTTCCCGC	773
Gly Ala Gly Pro *	
240	
TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTGCGAG	832

Figure 2

```
>sp|P08622|DNAJ_ECOLI DNAJ PROTEIN >pir||HHECDJ heat shock protein dnaJ -  
Escherichia coli >gi|145769 (M12565)-heat shock protein dnaJ  
[Escherichia coli] >gi|216441 (D10483) dnaJ protein [Escherichia  
coli]  
Length = 376
```

```
Score = 138 (63.7 bits), Expect = 1.2e-10, P = 1.2e-10  
Identities = 25/62 (40%), Positives = 39/62 (62%)
```

```
Query: 35 YYELGIVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRS 94  
      YYE+LGV A E+++A+ + + HPDR+ G+ ++F E+ EAY VL+ Q R +  
Sbjct: 6 YYEILGVSKTAEEREIRKAYKRLAMKYHPDRNQGDKEAEAKFKEIKEAYEVLTD SQKRAA 65  
  
Query: 95 YD 96  
      YD  
Sbjct: 66 YD 67
```

Figure 3

>gi|1703590 (U80439) contains similarity to a DNAJ-like domain [Caenorhabditis elegans]
Length = 345

Score = 98 (45.2 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12
Identities = 17/37 (45%), Positives = 28/37 (75%)

Query: 28 QRSRPSTYYELGVPHGASTEVEVKRAFFSKSKELHPD 64
++ R T+YE+LGV A+ E+K AF+++SK++HPD
Sbjct: 22 KKIRQORTHYEVLGVESTATLSEIKSAFYAQSKKVHPD 58

Score = 74 (34.1 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12
Identities = 17/32 (53%), Positives = 19/32 (59%)

Query: 71 SLHSRFVELSEAYRVLSREQSRRSYDDQLRSG 102
S + F+EL AY VL R RR YD QLR G
Sbjct: 64 SATASFLELKNAYDVLRRPADRLDYQLRGG 95

Score = 39 (18.0 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12
Identities = 10/42 (23%), Positives = 19/42 (45%)

Query: 162 LLMLAGMGLHYIAFRKVKQMHLNFMDEKDRIITAFYNEARAR 203
L+++AG Y+ Q L+ + ++D I F + R
Sbjct: 158 LVLVAGYNGGYLYLLAYNQKQLDKLIDEDEIAKCFLRQKEFR 199

>gnl|PID|e281266 (Z81030) C01G10.12 [Caenorhabditis elegans]
Length = 191

Score = 96 (44.3 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09
Identities = 17/41 (41%), Positives = 27/41 (65%)

Query: 35 YYELGVPHGASTEVEVKRAFFSKSKELHPDRDPGNPSLHSR 75
YYE++GV A+ +E++ AF K+K+LHPD+ + SR
Sbjct: 19 YYEIIIGVSASATRQEIRDAFLKTKQLHPDQSRKSSKSDSR 59

Score = 54 (24.9 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09
Identities = 10/22 (45%), Positives = 15/22 (68%)

Query: 75 RFVELSEAYRVLSREQSPFSYD 96
+F+ + EAY VL E+ R+ YD
Sbjct: 71 QFMLVKEAYDVLFPTEKPKKEYD 92

Score = 35 (16.1 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09
Identities = 9/44 (20%), Positives = 22/44 (50%)

Query: 141 QGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFRKVKQMHLN 184
+ P+ + KQ ++L ++A +G + + RK++ L+
Sbjct: 145 RNPEDEYLRKQKNRMLVVLAATVMALIGANTVYIRKLQADRLS 188

Figure 4

>sp|Q10209|YAY1_SCHPO HYPOTHETICAL 44.8 KD PROTEIN C4H3.01 IN CHROMOSOME I
>gi|1184014 (Z69380) unknown [Schizosaccharomyces pombe]
Length = 392

Score = 84 (38.8 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08
Identities = 13/35 (36%), Positives = 25/36 (69%)

Query: 35 YYELLGVHPGASTE EVKRAFFSKSKELHPDRDPGNP 70
YY+LLG+ A+ ++K+A+ + + HPD++P +P
Sbjct: 9 YYDLLGISTDATAVDIKKAYRKLAVKYHPDKNPDDP 44

Score = 64 (29.5 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08
Identities = 14/40 (35%), Positives = 23/40 (57%)

Query: 75 RFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRTTVHD 114
+F ++SEAY+VL E+ R YD + + P+ T +D
Sbjct: 50 KFQKISEAYQVLGDEKLRSQYDQFGKEKAVPEQGFTDAYD 89

Score = 37 (17.1 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08
Identities = 9/29 (31%), Positives = 15/29 (51%)

Query: 190 DRIITAFYNEARARARANRGILQQRQRL 218
DR A E A A+ + +++ RQR+
Sbjct: 149 DRKNAQIREREALAKREQEMIEDRRQRI 177

Score = 33 (15.2 bits), Expect = 0.00081, Sum P(3) = 0.00081
Identities = 8/19 (42%), Positives = 11/19 (57%)

Query: 140 PQGPQLRQQQHKQNKQVLG 158
PQG + Q+ + QVLG
Sbjct: 44 PQGASEKFQKISEAYQVLG 62

Figure 5

>gnl|PID|e253406 (X77635) tumorous imaginal discs [Drosophila virilis]
>gnl|PID|e263866 (Y07700) Tid58 protein [Drosophila virilis]
Length = 529

Score = 153 (70.6 bits), Expect = 9.7e-13, P = 9.7e-13
Identities = 27/71 (38%), Positives = 44/71 (61%)

Query: 26 AGQSRPSTYYELLGVHPGASTE EVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRV 85
+ R + YY L/GV A+ +++K+A++ +K+ HPD + +P +F ++SEAY V
Sbjct: 72 SSSRMQAKDYATLGVAKNANAKDIKKAYYELAKKYHPDTNKDDFDASKKFQDVSEAYEV 131

Query: 86 LSREQSRRSYD 96
LS +Q RR YD
Sbjct: 132 LSDDQKRREYD 142

Figure 6

MCG18	-----MPPLLPLRLCRLWP-RN--PP-----SRLLGAA
HDJ-2	MVKETTYDVLGVKPNATQEELKKAYRKLALKYHPDKN--PN----EGEKFKQISQAYEV
HDJ-1	MGKD--YYQTLGLARGASDEEIKRAYRRQALRYHPDKNKEPG----AEEKFKEIAEAYDV
HSJ1	M-AS--YYEILDVPRASADDIKKAYRRKALQWHPDKN--PDNKEFAEKKFKEVAEAYEV
* . . *	
MCG18	AQQRSRPSTY--YELGTVH-----PGA-----ST-EEVKRAFFS--
HDJ-2	LSDAKKRELYDKGGEQA IK-----EGGAGGG-----FGSPMDIFDMFFGGG
HDJ-1	LSDPKRREIFDRYGEEGLKGGSP-----SGSGGGANGTSF SYTFHGDPHAMFAEFTG--
HSJ1	LSDKHKREIYDRYGREGLTGTGTGPSRAEAGSGGP--G--FTFT-FK.SPEEVFREFFG--
* . . *	
MCG18	KSKELHPDRDPGNP---SLHSRFVELSEAYRVLSREQSRRS--YDDQLRSGSPPKSPRT
HDJ-2	GRMQRERRGKVVHQLSVTLEDLYNGATRKLALQKNVICDKCEGRGGKKGAVECCPNCRG
HDJ-1	GRNPFDITFFGQRNGEEGMDIDDPFSGFPMGGFTNVNFGRS--RSAQEPARKKQDPPVT
HSJ1	SGDPFAELFDDLGP--FSELQNRGSRHSGPFTTFSSSPGHSDFS SSSFSFSPGAGAFRS
* . . *	
MCG18	TVHDKSAHQTHSSWTPPNAQY---WSQFHSVRPQ-----GP-----QLRQQQHKQN
HDJ-2	TGMQIRIHQIGPMVQQIQSVCMECQGHGERISPK-DRCKSCNGRKIVREKKILEVHIDK
HDJ-1	HDLRVSLEEIYSGCTKMK-----ISH-KRLNP--D-----GKSIRNEDKILTIEVKK
HSJ1	VSTSTTFVQGRRIITRRIME-----NGQ-ERVEVEED-----GQ-----LKSVTINGVPD
* . . *	
MCG18	KQVLGYCLLL-----MLAGMGLHYIAFRKVKQMHLMFMDE-KDRIITAFYNearARARAN
HDJ-2	GMKDGQKITFHGEGDQEPGLEPGDIIIVLDQKDHAVFTRGEDLFMCMDIQLVEALCGFQ
HDJ-1	GWKEGTKITFPKEGDQTSNNIPADIVFVLKOKPHNIFKRDGSDVIYPARISLREALCGCT
HSJ1	DLARGLELSR-RE--QQP-SVTSRSGGTQVQQTASCPDLD-SDLSEDEDLQAMAYSLSE
* . . *	
MCG18	RGILQQRQLRQQRQPP-PSEPTQGPEIVPRGAGP-----
HDJ-2	KPISTLDNRTIVITSHPGQIVKHGDIKCVLNEGMPYRRPYEKGRLLIEFKVNFPENGFL
HDJ-1	VNVPTLDGRTIPVVFK--DVIRPGMRKVPGEGPLPKTPEKRGDLIEFEVIFPER--I
HSJ1	MEAAGKKPAGGREAQHR-RQGRPRPSTKIQA WGGP--RR--VRG--VKQPNVHPQR-RR
* . . *	
MCG18	-----
HDJ-2	SPDKLSLLEKLLPERKEVEETDEMDQVELVDFDPNQERRRHYNGEAYEDDEHHPRGGVQC
HDJ-1	PQTSRTVLEQVLPI-----
HSJ1	PLAASSSEHRAQPD-----LIQILTGGSDSLWEEKRGVS-----
* . . *	
MCG18	---
HDJ-2	QTS
HDJ-1	---
HSJ1	---

* = amino acid identity in all 4 proteins

. = conservative substitution

Figure 7

CAAGGAGCCTCTGCCTGCCCCTCGTCGTCATGCCGTCCCTGTTGCTCCAGCTGCCCCTGC 60
M P S L L L Q L P L 10
GCCTATGCCGGCTGTGGCCGCATAGCCTTTCCATCCGACTTCTCACAGCCGCCACAGGGC 120
R L C R L W P H S L S I R L L T A A T G 30
AGCGGTCTGTCCCTACTAATTACTATGAATTGTTGGGCGTGCATCCGGGTGCCAGCGCTG 180
Q R S V P T N Y Y E L L G V H P G A S A 50
AAGAGATTAAACGTGCTTTTTTTCACCAAGTCAAAGAGCTACACCCTGATCGAGACCCTG 240
E E I K R A F F T K S K E L H P D R D P 70
GGAACCCAGCCCTGCATAGCCGCTTTGTGGAGCTGAATGAGGCATATCGAGTGCTCAGTC 300
G N P A L H S R F V E L N E A Y R V L S 90
GTGAGGAAAGTCGTCGTAAGTATGACCACCAGCTGCATTACAGCCAGTCCTCCAAAGTCTT 360
R E E S R R N Y D H Q L H S A S P P K S 110
CAGGGAGCACAGCCGAGCCTAAGTATACGCAACAGACACACAGCAGCTCCTGGGAACCCC 420
S G S T A E P K Y T Q Q T H S S S W E P 130
CCAACGCTCAATACTGGGCCCAGTTCCACAGTGTGAGGCCGAGGGGCCGAGTCAAGGA 480
P N A Q Y W A Q F H S V R P Q G P E S R 150
AGCAGCAGCGTAAACACAACCAGCGGGTCCTGGGGTACTGCCTCCTGCTCATGGTGGCAG 540
K Q Q R K H N Q R V L G Y C L L L M V A 170
GCATGGGCCTGCACTATGTTGCCTTCAGGAAGCTGGAGCAGGTGCATCGCAGCTTCATGG 600
G M G L H Y V A F R K L E Q V H R S F M 190
ATGAAAAGGACCGGATCATTACAGCCATCTACAATGACACTCGGGGCCAGGGCCAGGGCCA 660
D E K D R I I T A I Y N D T R A R A R A 210
ACAGAGCCAGGATTACAGCAGGAGCGCCACGAGAGGCAGCAGCCTCGGGCAGAACCCTCCC 720
N R A R I Q Q E R H E R Q Q P R A E P S 230
TGCCTCCAGAAAGCTCCAGGATCATGCCCCAGGACACAAGCCCCTGAGAGGCTTAATAA 780
L P P E S S R I M P Q D T S P * 245
ATGGGACCTTCATTGGTCCTCTCCCTGCTGCCTGTCCAGAACTACACGTGCAATAAACTC 840
ATTTTCAG(A)n 849

Figure 8

human MCG18	MPPLL---PLRLCRLWPRNPPSRLLGAAAGQSRSPSTYYELLGVHPGASTEEVKRAFFSK
mouse MCG18	MPSLLLQLPLRLCRLWPHSL SIRLLTAATGQSRVPINYYELLGVHPGASAEELKRAFFTK
	*** ** *****. *** ** *****.*** ** *****.*** ** *****.***
human MCG18	SKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRTTVHDKSA
mouse MCG18	SKELHPDRDPGNPALHSRFVELNEAYRVLSREESRRNYDHQLHSASPPKSSGSTAEPKYT
	*****.***** *****.*** ** *****.*** ** *****.*** ** *****.***
man MCG18	HQTHSS-WTPPNAQYWSQFHSVRPQGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFR
mouse MCG18	QQTHSSSWEPFNAQYWAQFHSVRPQGPESRKQQRKHNRVLRGYCLLLMVAGMGLHYVAFR
	*****.***** *****.*** ** *****.*** ** *****.*** ** *****.***
human MCG18	KVKQMHLNFMDEKDRIITAFYNEARARARANRGILQQERQLGQRQPPPPSEPTQGPE---
mouse MCG18	KLEQVHRFSFMDEKDRIITAIYNDTRARARANRARIQQER---HERQQPRAEPSLPPESSR
	*.*** *****.*** *****.*** *****.*** *****.*** *****.***
human MCG18	IVPRGAGP
mouse MCG18	IMPQDTSP
	*.*** **

Figure 9

ttgaagtctagccccatcctggtccaatgcgctcttggtagcctcctttcccagctgccc 60
* S L A P S W S N A L L V A S F P S C P

gcccgcgcgcATGCCGCCCTTACTGCCCCCTGCGCCTGTGCCGGCTGTGGCCCCGCAACCC 120
P A A M P P L L P L R L C R L W P R N P>

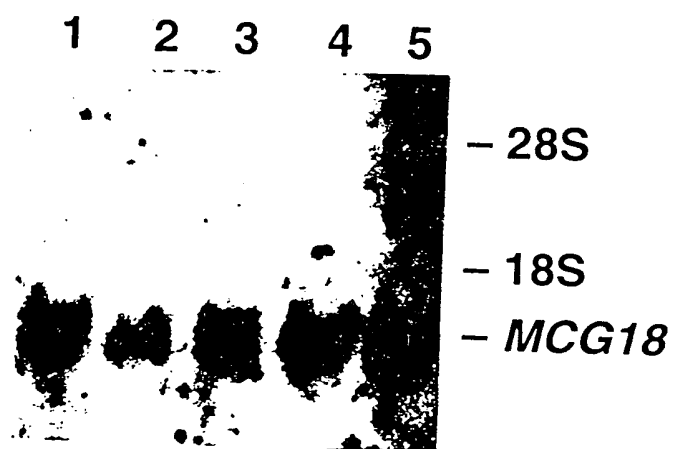


FIGURE 10